

TITLE OF THE INVENTION
SYSTEM AND METHOD FOR ENABLING INFORMATION ASSOCIATIONS

5 CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority under 35 USC §119(e) to provisional application serial number 60/146,469, entitled "System and Method for Enabling Information Associations", and filed July 30, 1999.

10

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR
DEVELOPMENT

N/A

15

BACKGROUND OF THE INVENTION

The present invention relates generally to systems for associating information objects together, and more specifically to a system and method for providing a "metalayer" of links and/or information for use in connection with the World Wide Web.

The growth of the World Wide Web ("Web"), generally thought of as those resources and users on the Internet using the Hypertext Transfer Protocol (HTTP), has been tremendous. Content information is organized on the Web in what are referred to as "Web pages." As it is generally known, a typical Web page contains a multiplicity of words, phrases, graphics, and other objects (henceforth called just "objects"), some of which are associated with what are known as "hyperlinks" or "links" to other Web pages. One of the powerful aspects of the Web is that users can go from one Web page to another, following links as they choose, in a

non-linear and unpredictable way. In existing systems, this is made possible by the designers of Web pages selectively assigning links to words, phrases, graphics, and other objects on their Web pages, for example through use of

5 "href" HTML tags within HTML files describing those Web pages.

The ability to follow links that are embedded within the HTML code describing a Web page defines the Web experience provided by existing systems. However, such an

10 approach may also hinder the user, in the sense that only certain objects within a given Web page, predefined by the designer of the Web page, are in fact linked to other Web pages. It would be desirable to shift some or all the responsibility for defining links for a given Web page from

15 the Web page designer, to a user who has selected that Web page, and who may wish to choose from a richer and more current selection of related links.

For example, consider the case of a Web user who is looking for information relating to a certain topic. Such a

20 user may be said to be "surfing" the Web. If the user has accessed a sports-related Web page, and is reading about a soccer player from Kosovo, the fact that the player is from Kosovo might trigger an association in the user's mind with the ongoing political unrest in that province of Yugoslavia.

25 However, the sports-related Web page probably does not have a pre-defined link associated with the word Kosovo, nor to any Web page that focuses on the political situation there. In existing systems, if the user wants to explore this new train of thought, he might have to leave the sports-related

30 Web page, go to a search engine, type in "Kosovo" as a search query, and thereby obtain a list of links related to Kosovo. Such a detour to a search engine could be time-

consuming and inconvenient, and the user might therefore instead opt to postpone and potentially forget about his sudden interest in Kosovo.

Existing application programs which seek to enhance a Web user's surfing experience include ThirdVoiceTM of ThirdVoice Inc., and products available from Hypernix Technologies, Ltd. ThirdVoice enables a user to post commentary relating to specific Web pages, without modifying the Web pages themselves, and without requiring permission of the Web sites at which the Web pages are located. ThirdVoice causes a caret to be displayed on top of existing content within a Web page, in order to indicate the presence of a message relating to that Web page. Users of ThirdVoice's software can see and read each other's posted messages. Hypernix enables users of its software to chat with other users visiting the same Web site, without the Web site's involvement. While Hypernix and ThirdVoice system provide groups of users with the ability to enter and retrieve comments regarding specific content within a Web page, they fail to change the underlying model of Web browsing, which is based on following sets of predetermined links from Web page to Web page.

For the above reasons, it would be desirable to have a system for using the Web which liberates Web users from the present Web surfing scenario, in which a Web surfer looking to go from one Web page to another is forced to choose a link from among the subset of objects on the page that have predefined links.

BRIEF SUMMARY OF THE INVENTION

In accordance with principles of the invention, a system and method are disclosed for providing a layer of "supercontent" above the existing layer of content on the Web, and which may be implemented by exploiting the use of client-side software running in tandem with a browser application program and a server component. The disclosed system and method enable information associations by parsing the objects contained within a document such as a Web page being downloaded by a browser program on a client system, in order to form a list of "metalinks" (also referred to as "metalink objects") within the document, for which related information may be available. The list of metalink objects may be formed by generating a unique object list for the document, and forwarding that list to a metadata server. The metadata server may then respond with a list of those objects within the list that are associated with related information. Those objects which are determined to have related information may then be considered metalinks. Alternatively, the client-side software may compare objects within the document with entries in a client-side data structure such as a dictionary, where each of the data structure entries represents at least one object for which related information is available. In general, objects for which related information is or may be available are referred to as "metalink" objects within the document.

The disclosed system may provide visual indications to a user of metalink objects within the document. Such visual indications of metalink objects may be provided automatically, or in response to a user action requesting such indications be made.

The disclosed system further operates to detect user selection of a metalink object in the document being displayed, for example when a user clicks on the display of a metalink object. In response to the user selecting a metalink object, the disclosed system initiates an object search by searching a database for information related to the selected metalink object. Such related information ("metadata") may include, for example, pointers or links to other documents. The related information is then displayed to the user, for example in a pop-up window referred to as the "metadata window." In response to user selection of one a data source indicated in the metadata window, the system causes the associated content (the "resource") to be retrieved and displayed, for example, in another window within the user interface of the local system.

The pop-up window used to display metadata is referred to herein as the "metadata window." A name of a piece of metadata displayed within the metadata window is referred to as a "data source", and the metadata itself is referred to as the associated "resource." Metadata may link directly to streaming audio/video or other multimedia content. When such metadata is selected by the user, the disclosed system may provide a separate window within which such multimedia content is presented to the user.

In another aspect of the disclosed system, the set of metadata provided to the user in response to selection of a metalink object is determined in part based on the user's past activities, the context within which the selection was made, or user profile information explicitly provided by the user to the system. In this regard, techniques such as Natural Language Processing (NLP) may be applied to determine contextual meaning associated with content that is

currently being displayed, or that has been previously selected by the user. Such contextual meaning may consist of themes or topics related to the selected content. For example, information related to metalink objects may be weighted in terms of their perceived relevancy to a user of the local system, or to a group of users spread across multiple systems, or to all users. In such an embodiment, the client software of the disclosed system monitors a number of activities on the local system which approximate or suggest the interest level of a user in a document that is currently being displayed. Such monitored activities may include time spent with the document open, whether the user scrolled through the document, whether the user clicked on a link within the document, and other user actions. A rating may then be generated for that document within a specific context, and applied to future requests for information related to a metalink object. The contextual analysis, user profiling, and/or explicitly entered user preferences may further be used by the disclosed system to perform word disambiguation with regard to selected metalink objects, in order to provide metadata related to the most likely intended meaning of a word having multiple potential meanings. A frame on one side of the current user display may further be provided by the disclosed system which displays a list of metalinks that relate to the current context, user preferences, and/or user profile information.

The disclosed system empowers a user to pursue tangential interests while surfing through various Web pages. Instead of forcing the user to choose from a small subset of the words or phrases within a document that happen to be conventionally "linkable" objects, the disclosed system potentially makes virtually every object on a Web

5

10

[illegible]

20

Fig. 1 is a block diagram showing an illustrative embodiment of the disclosed system;

25

30

Fig. 4 shows a Web page with illustrative metalinks underlined, in which a user has moved the cursor over one of

the metalinks, whereupon the cursor has been changed from an arrow into a modified "hyperlink hand" to indicate that "Stanley Kubrick" is a metalink;

Fig. 5 shows the Web page of Fig. 4 after the user has
5 clicked on the "Stanley Kubrick" metalink, whereupon a new window (the metadata window) has been generated containing information about related links;

Fig. 6 shows the Web page of Fig. 5 after the user has moved the mouse from the "Stanley Kubrick" metalink to the
10 window of related information (metadata); and

Fig. 7 shows the Web page of Fig. 6 after the user has clicked on the "Farewell, Stanley Kubrick" link within the metadata window, whereupon a new window has been generated showing the selected Web page.
15

DETAILED DESCRIPTION OF THE INVENTION

The disclosures of provisional application serial number 60/146,469, entitled "System and Method for Enabling
20 Information Associations", and filed July 30, 1999, to which this application claims priority under 35 USC §119(e), are hereby incorporated by reference herein.

As shown in the illustrative embodiment of Fig. 1, metalink associated client software 10 executes on the
25 client system 12, and operates to parse some number of objects in an HTML or Wireless Markup Language (WML) file 13 being downloaded through a browser program 14 also executing on the client system 12. Objects may consist of words, phrases, graphics or other display objects. For example, in
30 the case of graphics objects, the software 10 parses some or all associated text descriptors, for example as indicated by HTML "alt" tags in the HTML file 13. The client system 12

of Fig. 1 is further shown communicating over the World Wide Web 16 using HTTP 18 or WAP (Wireless Application Protocol). The client system 12 may include any hardware components suitable for providing a program code execution platform.

5 Such a program code execution platform may consist of, for example, a personal computer, personal digital assistant (PDA), workstation, server, wireless device, or Web appliance. It will also be recognized by those skilled in the art that such conventional hardware components may
10 include one or more processors for executing program code, a memory for storing program code for execution on the processors, and various input/output devices which may include secondary storage devices such as magnetic or optical disks and/or tape drives, and which may further
15 include user interface devices such as a keyboard, mouse, stylus, and display monitor.

During operation of the illustrative embodiment shown in Fig. 1, in order to access metadata associated with a metalink in HTML file 13, the user may initiate a request for an object search in a variety of ways. For example, the user may initiate a request for an object search with regard to a metalink by moving the mouse over the metalink and leaving it there for some predetermined amount of time, by clicking on the metalink, by hovering over the metalink, by single or double clicking on a metalink, by using the right mouse button to open a metalinks window, and/or by some combination of these and/or other techniques. Similarly, the user may move the mouse in an identifiable way (e.g., three quick loops) to signal that they are interested in seeing any available metadata for a given object. Along these lines, the user may identify an object of interest within the current Web page by dragging the cursor over it

with a mouse button depressed, thus causing it to be highlighted. Various other user interface mechanisms may be employed in connection with the disclosed system in order for the user to identify to the system one or more objects
5 for which related metadata is desired, as may be appropriate for specific implementations.

As a result of a user request for metadata, the disclosed system generates a window (referred to as the "metadata window") displaying a set of related information
10 (referred to as the "metadata"). The metadata within the metadata window may include the names of a number of pieces of metadata, referred to as "data sources." The actual pieces of metadata indicated by the data sources within the metadata window are referred to as "resources." The
15 metadata window may also display information that is unrelated to the metalink, such as weather reports or advertisements, in addition to the related metadata. The user can then move the mouse over a desired data source within the metadata window, and click on it, thereby gaining
20 access to the associated resource. For example, where a data source is a URL, clicking on the URL would direct the browser to download the HTML file describing the Web page indicated by the URL. Alternatively, some or all resources may be loaded automatically by the system, thereby not
25 requiring that the user explicitly load them. Other display rendering engines may also be used to download resources indicated by data sources other than URLs in the metadata window. While the metadata window may consist of a pop up window as described above, the metadata window may
30 alternatively consist of a side window or side frame within the user display.

0963330 "073400

Virtually any object may be a metalink. The disclosed system may be embodied, or may be configured by a user, such that all displayed objects are metalinks. In such an embodiment, the user may select any object that is displayed

5 to be used as an input to an object search for related information. However, the disclosed system may also be configured or embodied to provide "suggestions" of objects which the user may wish to perform object searches with. Such suggestions may be based on (1) user profile

10 information derived from past actions of the user, (2) user provided profile information, (3) user provided preferences, (4) contextual analysis of the currently displayed document, for example using natural language processing to extract themes contained in the current document or in a history of

15 recently loaded documents, (5) the presence of related information and/or metalinks associated with certain objects. When the disclosed system suggests objects to the user for which object searches may be desirable, it may do so by "marking up" the current document to visually indicate

20 the suggested objects. The specific style of markup may convey to the user information about the suggested object or its related information.

Various user actions may be employed to request that metalinks within the current display be visually indicated

25 to the user. For example, a technique referred to herein as a "white click" may employed for this purpose. In an embodiment of the disclosed system employing the white click activation technique, when the user performs a mouse click on some portion of the display that is devoid of content,

30 such as the margin of a document, the disclosed system will cause metalinks that are currently available to be visually indicated to the user. The disclosed white click technique

advantageously enables the user to activate visual indications of metalinks without having to look for a specific associated button within the user display. However, a button or menu, for example, could also serve
5 this function. The disclosed system may also be embodied using a right click to activate visual indication of currently displayed metalink objects, or using some combination of right click, white click, and/or other key strokes for this purpose. Additionally, the white click
10 activation technique, or some other activation technique, such as a right click or "glick" (a particular pattern of mouse movement, such as three clockwise circles), may further be used to bring up a search query window, into which the user may explicitly enter a search string to be
15 used by the disclosed system to search for related information.

Visual indications of metalink objects may include any kind of modification to the visual display of the metalink objects within a graphical user interface, including, but
20 not limited to, changes in format, color, font, or style of the metalink object displays. The appearance and selection of metalink objects within the user display is user configurable. Accordingly, the user may, for example, indicate through some type of menu interface how metalink
25 objects are to be visually indicated.

One technique for visual indication of metalink objects is through modification of the cursor image when the cursor is located relatively close to one or more metalink objects. Such changes in cursor image may be made in addition to or
30 instead of changes in the visual appearance of the metalink objects themselves, which also may be made when the cursor is moved over or near a metalink object. In one

illustrative embodiment, a "radius" of visual indication may be predefined or user configured. The radius of visual indication for a metalink object defines an area around a metalink which is "sensitized", such that when the cursor is moved by the user within the area, either the visual representation of the metalink object is modified to indicate that it is a metalink object, or the cursor itself is modified to indicate that it is located proximately to a metalink object. In this way, the visual indication of a metalink object may include modification of the display of the metalink object itself and/or modification of the cursor image.

In one such embodiment of the disclosed system, the user is notified that an object is a metalink object when client software causes the cursor to change to a predefined image, thus indicating the presence of available related information. For example, a special hand symbol, distinct from the hyperlink hand symbol used by existing browsers to indicate that an object is "hot", or linkable, may be employed for this purpose. Alternatively, the cursor may be modified to the hyperlink hand itself whenever it is over a metalink object. Moreover, an embodiment of the disclosed system may be provided which changes the style of an object in a Web page when metadata is or may be available for that object. Such style changing may include underlining and/or boldfacing of the object in a way that may or may not be distinctive from the default or page supplied styles for indicating hyperlinks.

Further, the user may expressly indicate to the disclosed system, for example through a menu type interface, a type or category of objects which should be visually indicated as metalinks. In this way, the user may configure

The markup style applied to any metalink object may depend on attributes of that object. For example, all objects which are associated with concepts of a specific category (e.g. sports) may be marked up a particular way. Alternatively or in addition, the markup style may indicate that the related metadata is particularly salient given the object context (including the context defined by a history of viewed Web pages), or the user's explicitly or implicitly established profile. Alternatively, or in addition, the markup style may indicate that the related metadata is of particularly high quality or quantity, etc.

ATTORNEY DOCKET NO. COMET-003XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & HAYES, LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

within the same document. Furthermore, the metadata available for a given object may be considered relatively more interesting to the user based on the amount of such related information, or based on individual or collective user profiles developed by the system in response to activities of a single user or a group of users. Such user profiles may include explicitly entered preferences or profile information provided by a user or users to the disclosed system, and/or the results of contextual analysis of currently or previously selected documents. Various contextual analysis techniques may be employed, for example based on text analysis algorithms and techniques. Such a relevancy determination by the disclosed system may be used to determine which ones of the currently displayed objects are to be visually indicated to the user as metalinks.

Alternatively, or in addition, the disclosed system may determine the specific set of metalink objects which are to be visually indicated through monitoring of actions performed by the user. In this way the disclosed system may provide "suggestions" to the user, based on the users behavior pattern, of which objects may be associated with further information related to the user's past browsing pattern. In the absence of any filtering, the disclosed system may provide visual indication to the user of all currently visible objects for which related information is available. In general, the visual indication of metalink objects may be used to inform the user that a relatively "rich" set of related information is available for some objects within the current display.

In one embodiment, as illustrated in Fig. 1, the disclosed system may include a dictionary of phrases, words, and other objects that are to be indicated as

metalinks to the user. Such metalink objects may or may not have metadata relating to them. In such an embodiment, the software 10 parses the HTML file 13 to identify those objects within the HTML file 13 matching entries within the dictionary 22. The software 10 may further operate to "highlight", in some predetermined manner, for example by boldfacing and/or underlining in a color-coded fashion, the displays of any objects matching an entry in the dictionary 22. Additionally, or alternatively, the disclosed system may detect when the user moves the mouse over or near one of these metalink objects, and as a result cause the cursor to be changed to a predetermined image, such as a modified hyperlink hand image, in order to indicate to the user that there are or may potentially be metadata related to the object over which the cursor has been moved. The modified hyperlink hand image may in fact be related in some way to the nature of the metalink object. For example, a symbol or word may be used as a cursor that is associated with the word or phrase that is the metalink object, such as a cursor including a corporate logo when the cursor is located over or near the name of the associated corporation.

The client software 10 of Fig. 1 further detects if the user clicks on any of the metalink objects shown on the Web page. In response to a user clicking on a display of a metalink object, the software 10 sends an object query message 30 to at least one metadata server, shown in Fig. 1 as metadata server 34 on the Web 16. The object query 30 contains or identifies the object that was clicked on, referred to as the "query object", and that is to be the subject of an "object search" performed on the metadata server 34. The metadata server 34 then returns a response message 32 including the results of the object search, i.e.,

a set of metadata related to the query object. The software 10 may cause a separate "metadata window" to be displayed by the client system 12. For example, the metadata window may be formed above or near the display of the query object. 5 Such a metadata window would contain a set of related metadata from among which the user could choose; clicking on any Data Sources within the metadata would send an instruction to the browser or other display device to open up the associated Web page or display the referenced 10 multimedia content or other information resource. The metadata window may also include other information and/or fields, such as news headlines, advertisements, search engine interfaces, etc. Such supplemental information within the metadata window may be related or unrelated to 15 the specific object query with which it is associated. For example, news headlines may be provided related to the object search query or results, whereas an advertisement may be provided based on settings provided by a user, or vice versa. The format of the metadata window may be partly or 20 completely configurable by the user, or may be predetermined by the system.

While the disclosed system may generate object searches in response to user selection of an object, an alternative embodiment may generate at least some object queries prior to a user selection of an object. Such unsolicited object queries may, for example, be performed by a background process, in order to prefetch object search results.

The disclosed system may employ a variety of approaches to generate related metadata for a given search object within the metadata server or servers. Embodiments of the disclosed system may interface to and/or inter-operate with existing databases that associate objects such as words

and/or phrases with specific URLs indicating the location of various types of content on the Internet. Such content may, for example, include HTML documents, graphics, video and/or audio, etc. Such existing databases include search engines, Web pages such as the Internet Movie Database (IMDB.COM), indexes, electronic news feed services, and directories for the Web. In the disclosed system, these databases can be used to supply metadata for metalinks. The disclosed system, running on the client and/or server system, can automatically supply a search input for these databases based on a search object selected by the user.

Embodiments of the disclosed system may utilize various techniques to improve the relevance of search results obtained from various databases. For example, existing search engines may not have information available to them regarding a specific user conducting a specific search. Accordingly, if ten different users conduct a search for the search query "mustang" on an existing search engine, they will likely each be provided the same search results. In contrast, and due to the ability of the client software to access, collect and interpret information regarding user behavior on the client system, possibly with the assistance of a server system used to process user profiles, the disclosed system can enrich information provided in the search query 30 shown in Fig. 1. For example, information relevant to an object search may reflect the location of content currently and/or previously selected by the user on client system 12, and/or any "theme" describing the purpose, nature or context of content stored on, or in association with, the client system 12. For example, words or phrases near or on the same page as the selected object, or within previously selected documents, can be used as a basis for

context analysis, for example using Natural Language Processing (NLP) techniques. The results of such context analysis can be used to form "guesses" regarding the user's intent and/or objectives with regard to an object search.

5 For example, these techniques might be employed to infer that when the user requests more information on the term "mustang", he is interested in the car rather than the mammal. This conclusion may be derived from the observation that the terms "Ford" and "car" also appear near the target
10 object. These correlations may be established, for example, by processing a large number of web pages which pertain the Ford Mustang. A statistical analysis, for example, of this set of web pages may reveal that the proximal appearance of the words "Ford" and "car" around the word "mustang"
15 strongly suggest the subject as that of the automobile.

Although the source of the query 30 is shown as a client computer system 12 in Fig. 1, this embodiment is shown only for purposes of illustration, and the present invention is not limited to such an embodiment.
20 Accordingly, the search query 30 used in connection with the disclosed system may be generated by any type of system, including a client system, server system, network appliance, or other, and which may generally be referred to in this regard as an "originating system" of the query 30. The
25 document from which the search query was generated may similarly be referred to as the "originating content" for the query. For example, if the originating content primarily or significantly stores information regarding horses, search results in the response 32 for a search query
30 30 including the term "mustang" might be skewed, either by software executing on the client or the server system, towards inclusion of other horse-related sites. On the

other hand, if the originating site is considered to be more car-related than horse related, the search results in the response 32 might be filtered or otherwise prioritized to emphasize sites relating to Ford Mustangs.

5 Additionally, embodiments of the disclosed system may include other features used to build "intelligence" into the metadata search process and thereby improve the relevance of the related metadata provided in the response 32. For example, using the software 10 executing on the client
10 system 12, the disclosed system can collect and analyze data that existing server-based search engines cannot access. In this regard, the software 10 may be employed to record how the user interacts with the Web page or pages associated with various metalinks. For example, the disclosed system
15 may be used to record whether a user of the client system 12 chooses to follow a given metalink from among a number of related metalinks. Additionally, the client software may further be employed to record the occurrence of certain predetermined actions of the user, such as:

- 20 1) spending a significant period of time at a given Web site or Web page,
 2) clicking on an existing link embedded within a Web site,
 3) performing an object search from a Web site,
25 4) scrolling down a particular page at a Web site, or
 5) performing any other action indicating a raised level of interest in a particular Web site.

 In this way, the disclosed system may determine that the user deemed a particular Web site, page, document or
30 piece of information relevant to his or her general interests, or relevant to the original search object, or relevant to the metalink from which it was reached. On the

other hand, certain actions performed by the user may indicate that the user did not find a particular Web site or page useful. For example, if the user immediately clicks the "back" button on the browser after a Web page is downloaded, or closes the window displaying that Web page, an embodiment of the disclosed system may determine that the user found the Web page relatively less relevant or even irrelevant. Messages describing the Web site or sites a user has found to be more or less relevant can be transmitted by the software 10 to the server system 34, and used to weight entries within the metadata database on the server system 34. Such information may be used, either alternatively or in addition, to perform such weighting on entries of a metadata database stored within the client system. In general, the metadata database may be stored within the server and/or client system. Entries may be weighted with respect to an individual user, a subset of all users deemed to be similar, or all users of the system. Such information may also be used to modify specific search queries generated on the server system 34 in response to object search queries received from the client system 12. With regard to an individual user, such search result weighting may be performed with or without explicit permission from that user, and could be started and stopped explicitly by the user, in order to provide user control over when and where the relevancy analysis takes place.

In this way the disclosed system can be requested by a user to develop and store a "user profile" that applies intelligence on an individualized basis to the search process, for the benefit of the requesting user. Thus, a sports fan who frequently chooses sports-related links whenever he does object searches can be shown more sports-

Information regarding user activity may also be collected anonymously by the disclosed system for activities

In one illustrative embodiment, the client software 10 further includes an Artificial Intelligence ("AI") engine 42, which learns to interpret content based on previous documents visited by each user. In this way, the disclosed system may predict the behavior and wishes of a specific user sufficiently to offer intelligent "guesses" as to the kinds of information in which the user is interested. This relevancy information may then be fed back to a database on the metadata server, and used to modify the contents of the database and/or provide weightings or modifications to be applied to specific search objects.

ATTORNEY DOCKET NO. COMET-003XX
WEINGARTEN, SCHURGIN,
GAGNEBIN & HAYES, LLP
TEL. (617) 542-2290
FAX. (617) 451-0313

the client system 12. In contrast, collection of user-behavior data would be difficult, if not impossible, for existing Web-server based search engines. Embodiments of the present system may exploit the fact that the software 10 executes on the user's "desktop", and can thus detect events indicating relevant user actions, such as where (or even whether) the user moves his mouse around on a current Web page, whether he clicks on a link or performs another object search on the current Web page, scrolls down the current Web page, spends a great deal of time on a given Web page, etc. Similarly, this analysis can be extended to cover user interaction with other software systems and data resident on the client system. This behavioral feedback data can be aggregated, interpreted, and applied in powerful and 15 unprecedented ways to improve future searches. Thus it is shown that, in order to optimize the efficacy of searches for related metadata, a combination of aggregated feedback for all users, user behavior profiles, and origination site information can advantageously be used to provide related 20 metadata.

The database(s) used on the server system 34 to store and access related metadata resulting from object searches may also be loaded based on submissions of Web site information from the publishers of Web sites. For example, 25 related metadata may be loaded into such databases in association with specific URLs or other search objects at the request of publishers and/or hosts of the Web sites indicated by such metadata, or related to such search objects. In addition to installing related information from 30 participating Web sites, the metalink object list itself can be supplemented by those Web sites. Users may visit such a Web site and be induced to "install" the set of related

information from that website. From that point on, that user's metalink highlighting may include those terms installed from said Web site, and information related to those terms will include that published by the same Web site.

In one embodiment, the disclosed system combines the above mentioned user-behavior feedback with a capability to receive URLs from users that are intended to "challenge" the metadata automatically generated by the disclosed system in response to an object search for a given search object. For example, if a Web site contains resources (e.g. web pages or URLs) primarily related to gardening, the publisher of that site may submit it to the disclosed system with the desire that users performing object searches on some predetermined gardening related objects (such as "roses") will be displayed a link to that site as part of the related metadata. In an illustrative embodiment, the disclosed system may provide a "challenger" portion of the metadata window presenting the names of challenger resources associated with an object search. Such a "challenger" list may include some number of challenger resources provided by Web site providers, that are "challengers" to the main or "incumbent" metadata provided by the object search. The user may then decide whether to follow one of the ordinary metadata resources selected by the system or one of the "challenger" resources. For example, the aforementioned Web site pertaining to gardening may appear as a "challenger" metadata resource, meaning that any time a user performs an object search on roses, that site will appear in a special "challenger" category displayed within the metadata window. The disclosed system subsequently employs user-behavior feedback and/or explicit feedback (e.g., voting), to rate

all metadata, including challenger metadata. If a "challenger" Web site performs relatively well, in the sense that the disclosed system detects a sufficient level of user activities indicating it is often relevant to user needs, then the challenger site may become "weighted" in the metadata database to such an extent that it appears in the main metadata results for the "rose" search object.

Placement of a Web site, and/or relative ordering of Web sites within search results for a given set of one or more search objects may also be predetermined in response to payments made for specific placements and/or ordering. For example, when users search on the word "car", various car-related Web sites appear as metadata within the metadata window. A particular car-related site, however, could purchase the right to have its Web site's URL automatically appear among the related metadata for the object "Cadillac", even if that Web site would not normally be selected for display to the user. This would enable companies to "purchase" certain keywords as search objects, and thus ensure that their Web site appears whenever someone does a related object search. In other words, the amount of money paid by such a company could be used to "weight" that company's metadata in the system's decision about which metadata to display to the user for a given object search. In one embodiment, the disclosed system allows real-time communication between a current Web site and the client system, such that the Web site could dynamically control the behavior of the client system. For example, the Web site could ask that a predetermined icon with a predetermined meaning, such as a red star meaning "special information", appear next to a specific metadata entry, or it could

suppress the appearance of a competitor's link within the metadata.

In one embodiment, the client software 10 is automatically started when the client system 12 boots, and is capable of injecting itself into the address space of various software applications on the client system 12, including the Web browser 14. Examples of the Web browser 14 include Internet Explorer provided by Microsoft Corporation, as well as Netscape Navigator provided by Netscape Corporation.

The client software 10 operates to intercept the internal operations of the browser 14. For example, the client software may monitor the content of a Web page as it is downloaded by the browser 14 for display to the user. Such monitoring may, for example, be accomplished through an Application Program Interface (API) provided by the browser 14, or through various other techniques. For example, the client software 10 may read the HTML document source text describing a Web page being downloaded by the browser 14 by intercepting network packets received by the client system 12, or by intercepting read operations generated by the browser 14 to a local file system on the client system 12, for example when loading the HTML document from a local cache. The client software 10 may also be embodied to intercept various user inputs to the browser 14, for example by acting as a filter for mouse movement, mouse button and keyboard input. Windows related messages and/or operating system (OS) system calls may also be processed by the client software 10 before being relayed to the browser 14 through a variety of intercepting techniques, including, but not limited to, a combination of application and operating system component intercommunication interception.

Further during operation of the client software 10, an internal representation 40 is generated, representing the structure of a downloaded Web page associated with the HTML file 13, including image data, textual content, metalink objects within the Web page, etc. The client software 10 further operates to differentiate content from layout information and extract text from within the downloaded Web page, and may use probabilistic/statistical natural language processing algorithms to build a set of contextual information from the page.

The client or server builds and stores a database or "dictionary", for example either locally in the computer's memory, or remotely on a server system, of information that it may use to specify object-metadata associations. For example, as shown in Fig. 1, the client software 10 may generate the dictionary 22, storing possible keywords or other objects which may be used as inputs to "object searches." The client can parse the text and other objects on the page and determine which objects (phrases, words, e.g.,) should be indicated as metalinks. It can accomplish this by analyzing these objects along with the database and/or dictionary, possibly employing natural language processing and related algorithms to "make sense of"/"disambiguate" the objects on the page. The dictionary may be augmented and updated in response to such analysis through direct or indirect user input and/or via communications with a networked server. To reduce latency associated with network communication, and/or the amount of data transmitted across the network, the dictionary may be partially or totally stored on the client. To allow for the object database and metadata associations to change over time, any such client dictionary could be periodically

5

15

30

disclosed system, and/or through other search engines. The user's selection determines what keywords and/or search query terms are employed by the disclosed system to obtain any related information.

5 Once a search object has been selected by the user on
the client system 12, the client software 10 may contact the
server system 34, which initiates a search for relevant
information. The client software 10 may use HTTP via its
own pool of network connections. The client software 10
10 may, for example, interrogate the browser 14 for any
relevant configuration information regarding proxy servers,
so that the client software 10 is able to successfully
complete an HTTP request to the Internet. Although the
communications protocol employed by the client software 10
15 in an illustrative embodiment may be HTTP, the returned
document may be HTML, or XML, or some other kind of
proprietary text or binary format, as appropriate to the
specific application.

The server 34 may return an HTTP response that comprises an index of related pages and/or topics. This data is machine readable and is to be interpreted by the client software 10 in order to generate the metadata presented to the user on the client system 12. The client software 10 interprets the data in the HTTP response, and creates a metadata GUI for the user on the client system 12. The metadata GUI may be a popup menu, or a pop-up window containing a graphical list of search results that the user may select with the keyboard or mouse in order to load an associated page onto the client system 12.

30 The server system 34 may form topics that reflect
current research topics that are being investigated by
significant groups of users. As multiple search queries

5
10
15
20
25
30

dynamically to provide metadata which reflect recent activities of users, such as large numbers of searches related to a breaking news report.

As mentioned above, the client software 10 is able to
5 cause an entry in the search engine database on the server system 34 to be "weighted" to reflect its usefulness based on (a) contextual data extracted from other pages downloaded onto the client system 12, (b) the amount of time that a page associated with the entry was viewed by the user on the
10 client system 12, (c) repeat visits to the page associated with the entry, and other criteria. The client software 10 may further also employ techniques to

1) determine whether and how the mouse is being moved over a page in order to infer either that the page is being
15 actively read, or, in the case where the mouse is inactive for some period of time, that the user has possibly walked away from computer;

2) determine whether the page is in the currently active window or has been covered by another window;

20 3) determine whether the page has been minimized; and

4) determine whether the user is scrolling the page.

Other techniques for determining the perceived relevance of a given page may also be employed. For
25 example, in the case of a hand held wireless device, such as a personal digital assistant (PDA), the disclosed client software may monitor such events as whether stylus is currently being used. The client is able to intercept this information because it has injected itself into the
30 browser's address space and can intercept the browser's function calls to the operating system, the browser's network communications, and disk activity.

096E8880 "073100

The client software 10 may, for example, run as its own process, and therefore be capable of intercepting the activities of many browsers executing on the client system 10. Similarly, the client software 10 may also operate to
5 intercept activities of other applications executing on the client system using similar techniques. Accordingly, the client software 10 may operate to intercept the activities of email clients, word processors, and other applications
10 executing on the client system 12, all of which may act as input to the search system, in order to provide contextual weighting for entries in the metadata search database, and/or as a basis for structuring subsequent object searches themselves.

In an illustrative embodiment, an interface may be
15 provided to the disclosed system which permits the metadata search database to be "seeded" by an administrator. Through such an interface, an "editorial view" may be imposed on the metadata search results, and a number of topics explicitly created by a system administrator, rather than by the
20 relatively slower process of analyzing user habits. For example, when a major news story breaks, a system administrator might create a topic containing keywords for that story. The topic would further be associated with a set of metadata, which would be provided in response to an
25 object search including any of the keywords for the story.

Fig. 2 shows a number of steps performed by an illustrative embodiment of the disclosed system. As shown in Fig. 2, at step 60 the client software is installed on the client system, such as a personal computer, and
30 configured to start up at system boot time and run in the background. At step 62, in response to a boot of the client system or some other trigger event, the client software

interacts with a programming interface provided by, or injects itself into the execution context of, an application program also executing on the client system, for example a browser program, so as to provide a method of interprocess
5 communication or interaction between the disclosed system and the host application. At step 64, the client software monitors actions performed by the browser, for example by intercepting browser related actions such as browser function calls, browser related network traffic, and/or
10 browser disk accesses.

At step 66, the client software reacts to the downloading of a Web page by the browser, for example by parsing text and other data contained within the Web page. This parsing is performed in order to identify which content
15 objects within the page are potentially metalink objects, and also to determine context information with regard to the page.

At step 68, the client software builds a list of metalink objects, such as keywords, phrases, and images,
20 contained within the downloaded Web page that should be indicated as metalinks to the user, and also the markup style to be applied to each object (if any). Such a list of metalink objects may be generated in a variety of ways. For example, the client software may generate a unique word list
25 based on the contents of the downloaded Web page, and then send that list to the metadata server. The metadata server may then provide a list indicating the words within the unique word list for which metadata is available through the metadata server. Alternatively, the client software may
30 automatically select a number of metalink objects within a Web page when the Web page is downloaded, in response to a Natural Language Processing (NLP) subsystem that "reads" the

text of the downloaded Web page and picks out significant subject matter and proper nouns, and compares it to a local database of such subject matter. Alternatively, the client software may just send the URL of the target page, along
5 with contextual information supplied by the user's browser (e.g. cookies, user agent, HTTP header information), so that the server system can independently retrieve and analyze the page. Such automatic generation of a list of metalink objects by the client system may also be provided in
10 response to the client software operating to break down the text of the downloaded Web page into tokens, and then transmitting those tokens to a metadata server that looks each one up in a database to determine whether any associated metadata exists.

15 In a further alternative embodiment, the initial list of metalink objects may be generated by comparing objects within the downloaded Web page to entries in a dictionary of potential metalink objects stored on the client system. While an initial list of metalink objects may be generated
20 on the client system without reference to the metadata server system, the client system may subsequently send the initial set of metalink objects for the Web page to the metadata server, which may then analyze the metalink object list for contextual meaning, and/or to determine whether any
25 metadata in fact exists for each of the metalink objects in the list. Indication of whether metadata exists may be provided by the metadata server to the client system, which may then modify the list of metalink objects for the Web page to only include objects for which metadata definitely
30 exists. Such an implementation would allow the data storage responsibilities of the disclosed system to be shared between the client and server, and also for the division of

10

20

25

user has selected a URL from within the metadata displayed to the user at step 76, for example by clicking on the URL in the generated metadata window. The content associated with the selected metadata may then either be displayed to the user through the client software at step 79, or, in the case where the selected metadata indicates a Web page, downloaded to the client system through the browser.

Fig. 4 shows a Web page with illustrative metalink objects 100 indicated to a user using a distinctive visual indication, in this case, underlining. For purposes of illustration, the distinctive visual indication of the metalink objects 100 is shown as underlining. Note that the invention does not require that metalink objects be distinctively identified. Such identification (or the lack thereof), and the meaning of any particular markup style, could be configurable by the user or defined by the software system. Indeed, it is possible that literally every object on the page is a metalink. The underlining of metalink objects as shown in Fig. 4 may employ a distinctive color of underlining, such as yellow. Note that the distinction attributed to the object may vary based on the type of object, what the object represents, or what sort of related information is available for the object. Again, this behavior may be specified by the user, or defined by the software system. The visual indication of the metalink objects 100 may further be provided, alternatively or in addition, by modification of the cursor when the cursor is passed over or near a metalink object. Alternatively, the visual indication may be explicitly activated by the user. For example, by clicking the right mouse button of the mouse, or clicking a button placed in a menu of the browser.

As also shown in Fig. 4, when a user has moved the cursor

25 The invention is not limited in application to content
provided through Web pages. The disclosed system may be
applied to email documents, for example. For example, with
existing systems, if a first user wants to tell a second
user about a new company, the first user might send the
30 second user an email describing the company along with the
URL of the company's Web site. With the disclosed system,
rather than the first user typing in the URL of the

company's Web site, if the invention software is running on the second user's computer, the second user can simply conduct an object search on the name of the company. Among the related metadata that appear might be the company's Web site. Other metadata might also appear, providing the second user with a rich choice of metadata resources should he want to get more information about that company. The present invention may be applied beyond the Web, and beyond email, to any applications that employ parsable content (i.e., text, video, graphics, etc.), such as word processors, spreadsheet programs, as well as applications which process multimedia content, and in which users may desire to see information that is related to a given object within the parsable content.

Those skilled in the art should readily appreciate that the programs defining the functions of the present invention can be delivered to a computer in many forms; including, but not limited to: (a) information permanently stored on non-writable storage media (e.g. read only memory devices within a computer such as ROM or CD-ROM disks readable by a computer I/O attachment); (b) information alterably stored on writable storage media (e.g. floppy disks and hard drives); or (c) information conveyed to a computer through communication media for example using baseband signaling or broadband signaling techniques, including carrier wave signaling techniques, such as over computer or telephone networks via a modem. In addition, while the invention may be embodied in computer software, the functions necessary to implement the invention may alternatively be embodied in part or in whole using hardware components such as Application Specific Integrated Circuits or other hardware, or some combination of hardware components and software.

While the invention is described through the above exemplary embodiments, it will be understood by those of ordinary skill in the art that modification to and variation of the illustrated embodiments may be made without departing from the inventive concepts herein disclosed. Specifically, while the preferred embodiments are disclosed with reference to use with a browser application program, the present invention is generally applicable to any type of program provided to a computer system. Moreover, while the preferred embodiments are described in connection with various illustrative data structures, one skilled in the art will recognize that the system may be embodied using a variety of specific data structures. Accordingly, the invention should not be viewed as limited except by the scope and spirit of the appended claims.